

The Origin and Evolution of the Galaxy Mass–Metallicity Relation

Xiangcheng Ma,¹ Philip F. Hopkins,¹ Claude-André Faucher-Giguère,² Nick Zolman,¹ Alexander L. Muratov,³ Dušan Kereš³ and Eliot Quataert⁴

¹ TAPIR, MC 350-17, California Institute of Technology, Pasadena, CA 91125, USA

² Department of Physics and Astronomy and CIERA, Northwestern University, 2145 Sheridan Road, Evanston, IL 60208, USA

³ Department of Physics, Center for Astrophysics and Space Sciences, University of California at San Diego, 9500 Gilman Drive, La Jolla, CA 92093

⁴ Department of Astronomy and Theoretical Astrophysics Center, University of California Berkeley, Berkeley, CA 94720

Draft version 10 April 2015

ABSTRACT

We use high-resolution cosmological zoom-in simulations from the Feedback in Realistic Environment (FIRE) project to study the galaxy mass–metallicity relations (MZR) from $z = 0$ – 6 . These simulations include explicit models of the multi-phase ISM, star formation, and stellar feedback. The simulations cover halo masses $M_{\text{halo}} = 10^9$ – $10^{13} M_{\odot}$ and stellar masses $M_* = 10^4$ – $10^{11} M_{\odot}$ at $z = 0$ and have been shown to produce many observed galaxy properties from $z = 0$ – 6 . For the first time, our simulations agree reasonably well with the observed mass–metallicity relations at $z = 0$ – 3 for a broad range of galaxy masses. We predict the evolution of the MZR from $z = 0$ – 6 , as $\log(Z_{\text{gas}}/Z_{\odot}) = 12 + \log(\text{O}/\text{H}) - 9.0 = 0.35 [\log(M_*/M_{\odot}) - 10] + 0.93 \exp(-0.43z) - 1.05$ and $\log(Z_*/Z_{\odot}) = [\text{Fe}/\text{H}] + 0.2 = 0.40 [\log(M_*/M_{\odot}) - 10] + 0.67 \exp(-0.50z) - 1.04$, for gas-phase and stellar metallicity, respectively. Our simulations suggest that the evolution of MZR is associated with the evolution of stellar/gas mass fractions at different redshifts, indicating the existence of a universal metallicity relation between stellar mass, gas mass, and metallicities. In our simulations, galaxies above $M_* = 10^6 M_{\odot}$ are able to retain a large fraction of their metals inside the halo, because metal-rich winds fail to escape completely and are recycled into the galaxy. This resolves a long-standing discrepancy between “sub-grid” wind models (and semi-analytic models) and observations, where common sub-grid models *cannot* simultaneously reproduce the MZR and the stellar mass functions.

Key words: galaxies: formation – galaxies: evolution – cosmology: theory

1 INTRODUCTION

The galaxy mass–metallicity relation (MZR) is one of the most fundamental properties observed in galaxies. In the local universe, there is a tight correlation between galaxy stellar mass and gas-phase oxygen abundance for star-forming galaxies (e.g. Tremonti et al. 2004), with an intrinsic scatter of only 0.1 dex in $\log(\text{O}/\text{H})$. This relation has been extended to local dwarf galaxies and found to be a uniform, tight correlation over five orders of magnitude in stellar mass, from $M_* = 10^6$ – $10^{11} M_{\odot}$ (Lee et al. 2006). Many different groups have confirmed the MZR exists at least to $z \sim 2.3$ (Zahid et al. 2011; Yabe et al. 2014; Erb et al. 2006; Sanders et al. 2014; Steidel et al. 2014). Zahid et al. (2013) compiled all the observed MZR at different redshifts and found the MZR smoothly evolves from $z = 0$ – 2.3 , with higher metallicity at lower redshift for a given stellar mass. The MZR is also found beyond $z > 3$ (e.g. Maiolino et al. 2008; Mannucci et al. 2009), despite the fact that the samples are still quite small.

Gas-phase metallicities represent the “current” state of chemical enrichment in the galaxies, while stellar metallicities reflect the “time-averaged” galactic metallicity across the whole star for-

mation history. Similarly, an MZR is also found in stellar metallicities. Gallazzi et al. (2005) derived the stellar metallicities for $\sim 44,000$ galaxies from SDSS and found a tight correlation between stellar mass and stellar metallicity for galaxies of stellar masses 10^9 – $10^{12} M_{\odot}$. Kirby et al. (2013) measured the metallicities of individual stars in a sample of dwarf galaxies within the Local Group and found the SDSS stellar MZR can be continually extended down to $10^3 M_{\odot}$. Despite the fact that stellar metallicity is challenging to measure at high redshifts, the stellar MZR provides very important and complimentary insights on the chemical evolution of galaxies, especially for massive quiescent galaxies and satellite galaxies in the local group where the gas-phase metallicities are hard to measure due to their low gas content.

Simple analytic models of galactic chemical evolution, such as the “closed box”, “leaky box”, or “accreting box” models (e.g. Schmidt 1963; Talbot & Arnett 1971; Searle & Sargent 1972; Edmunds 1990), are often quoted to explain the qualitative behavior of the MZR. Such models have also been modified to work in cosmological contexts and describe the evolution of MZR (e.g. Finlator & Davé 2008; Davé et al. 2012; Lilly et al. 2013). These indicate

that the observed MZR is the consequence of an interplay between star formation efficiency, metal loss from gas outflows, and gas recycling and accretion. For example, the stellar mass–halo mass relation (e.g. Moster et al. 2013; Behroozi et al. 2013) indicates that the star formation efficiency (fraction of baryons turned into stars) is lower in low-mass galaxies than in more massive galaxies, suggesting that low-mass galaxies should be less metal-enriched. Meanwhile, galactic winds are ubiquitous (see e.g. Veilleux et al. 2005, for a recent review), carrying metals away from galaxies. Low mass galaxies have shallow potential wells so they tend to lose a significant fraction of their gas and metals, while massive galaxies have potential wells deep enough to prevent material from escaping the galaxy (e.g. Dekel & Silk 1986). On the other hand, gas inflows bring the metal-poor gas in the galactic halo and/or in the intergalactic medium (IGM) inwards, diluting the metal content in the interstellar medium (ISM) and supplying new material for star formation (e.g. Kereš et al. 2005; Faucher-Giguère et al. 2011). During this process, a considerable fraction of the gas and metals that have been formerly ejected via outflows eventually come back to the galaxy (e.g. Bertone et al. 2007; Oppenheimer et al. 2010). Galaxy mergers and AGN activities could also be important, in the sense that they can trigger violent starburst, drive intensive gas outflows, and ultimately quench the star formation in the galaxy (e.g. Springel et al. 2005; Hopkins et al. 2013a).

These processes are tightly connected to each other and must be treated self-consistently to understand the complete picture of galactic chemical evolution. Semi-analytic models (SAMs) of galaxy formation include physically and/or empirically motivated prescriptions of heating and cooling, star formation, metal enrichment, gas outflows, accretion, and AGN feedback (e.g. Croton et al. 2006; Somerville et al. 2008; Benson 2012; Lu et al. 2014). They are able to trace the evolution of galaxies for a broad range of masses and predict many galaxy properties at different redshifts. However, one of the major challenges for SAMs is the difficulty in getting the stellar mass, star formation rate, and the metallicity for low-mass galaxies *simultaneously* consistent with observations. In particular, the metallicities of low-mass galaxies are very sensitive to the prescription of outflows. Even if different models can be fine-tuned to give reasonable stellar masses at $z = 0$, their predictions generally have large discrepancies with the observed MZR at high redshifts (e.g. Lu et al. 2014, for a detailed comparison and discussion).

More sophisticated studies of galaxy chemical evolution rely on cosmological hydrodynamic simulations of galaxy formation (e.g. Bertone et al. 2007; Davé et al. 2011b; Torrey et al. 2014). However, most of these simulations adopted very approximate, or “sub-grid” model of ISM physics, star formation, and stellar feedback, and use empirical prescriptions for galactic winds. Simulations with such simple treatments reveal very similar problems to the SAMs. For example, Torrey et al. (2014) found a very steep slope (much steeper than observed) at the low-mass end of their MZR, and attributed it to the low metal retention efficiency in the presence of strong outflows, which were required in their model in order to prevent low-mass galaxies from forming too many stars. These authors emphasized the tension between suppressing star formation and retaining enough metals in low-mass galaxies — this is exactly the same problem in some SAMs (see Lu et al. 2014). Moreover, the star formation histories in these simulations are very different and not all consistent with observations at high redshifts. Many cosmological simulations tend to form *too many* stars at early times (e.g. Davé et al. 2011a; Sparre et al. 2015; Fiacconi et al. 2015; for a review, see Somerville & Davé 2014). Such problems

is likely a result of imperfect star formation and stellar feedback models implemented in those simulations (see also Scannapieco et al. 2012, for a discussion). As a consequence, the evolution of the MZR is very different among these simulations.

Therefore, when using cosmological hydrodynamic simulations to understand the behavior of the MZR and its evolution, one must “correctly” capture star formation, stellar feedback, and gas inflows and outflows over the whole cosmic time. In this work, the first of a series, we will study the chemical evolution of galaxies using the FIRE (Feedback in Realistic Environment) simulations (Hopkins et al. 2014). The FIRE project¹ is a series of cosmological zoom-in simulations that are able to follow galaxy merger history, interactions of galaxies with the IGM, and many other important processes. These simulations include a full set of realistic physics and explicit resolution of the multi-phase ISM, star formation, and stellar feedback, *without* adjusted parameters. The FIRE simulations successfully reproduce many observed galaxy properties, including the stellar mass–halo mass relation, star formation histories, Kennicutt-Schmidt law, and the star-forming main sequence, from $z = 0$ –6, for a broad range of galaxy masses in $M_* = 10^4$ – $10^{11} M_\odot$ (Hopkins et al. 2014). Also, the FIRE simulations predict reasonable covering fractions of neutral hydrogen in the halos of $z = 2$ –3 Lyman Break Galaxies (LBGs Faucher-Giguère et al. 2015) and self-consistently generate galactic winds with velocities and mass loading factors broadly consistent with observational requirements (Muratov et al. 2015). These results further justify their reliability to study the chemical evolution of galaxies.

This paper focuses on the galaxy mass–metallicity relation. In companion papers, we will also study the stellar metallicity distribution functions and $[\alpha/\text{Fe}]$ abundance ratio variation in dwarf galaxies, metallicity gradients and their origins, metal outflows and recycling. We start by describing the simulations in Section 2 and present the mass–metallicity relation at different redshifts in Section 3. In Section 4, we discuss the processes that affect the galactic metal abundance. We summarize and conclude in Section 5.

2 THE SIMULATIONS

2.1 Simulation Details

This work is part of the FIRE project. A full description of the numerical methods and physics included in our simulations is presented in (Hopkins et al. 2014, and references therein). We summarized their main features here. All the simulations use the newly developed GIZMO code (Hopkins 2014) in “P-SPH” mode. P-SPH adopts a Lagrangian pressure-entropy formulation of the smoothed particle hydrodynamics (SPH) equations (Hopkins 2013), which eliminates the major differences between SPH, moving-mesh, and grid codes, and resolves many well-known issues in traditional density-based SPH formulations. The gravity solver is a heavily modified version of the GADGET-3 code (Springel 2005); and P-SPH also includes substantial improvements in the artificial viscosity, entropy diffusion, adaptive time-stepping, smoothing kernel, and gravitational softening algorithm.

We use the multi-scale “zoom-in” initial conditions generated with the MUSIC code (Hahn & Abel 2011), using second-order Lagrangian perturbation theory. The first set of simulations have been run down to $z = 0$ and cover halo masses 10^9 – $10^{13} M_\odot$ and stellar masses 10^4 – $10^{11} M_\odot$ at $z = 0$ (**mx** series). Most of them have been

¹ FIRE project website: <http://fire.northwestern.edu>

Table 1. Simulation Initial Conditions.

Name	M_{halo} (M_{\odot})	m_b (M_{\odot})	ϵ_b (pc)	m_{dm} (M_{\odot})	ϵ_{dm} (pc)	Merger History	Notes
m09	2.5e9	2.6e2	1.4	1.3e3	30	normal	isolated dwarf
m10	0.8e10	2.6e2	3.0	1.3e3	30	normal	isolated dwarf
m10lr	0.8e10	2.1e3	2.1	1.0e4	35	normal	low resolution
m10v	0.8e10	2.1e3	7.0	1.0e4	70	violent	–
m11	1.4e11	7.0e3	7.0	3.5e4	70	quiescent	–
m11v	3.3e11	5.6e4	7.0	3.0e5	140	violent	–
m12v	6.3e11	3.9e4	10	2.0e5	140	violent	several $z < 2$ mergers
m12q	1.2e12	7.1e3	10	2.8e5	140	late merger	–
m12i	1.1e12	5.0e4	14	2.8e5	140	normal	large ($\sim 10 R_{\text{vir}}$) box
m13	6.0e12	3.6e5	21	2.2e6	210	normal	small group mass
z2h350	7.9e11	5.9e4	9	2.9e5	143	normal	–
z2h400	7.9e11	5.9e4	9	2.9e5	143	quiescent	–
z2h450	8.7e11	5.9e4	9	2.9e5	143	normal	–
z2h506	1.2e12	5.9e4	9	2.9e5	143	violent	–
z2h550	1.9e11	5.9e4	9	2.9e5	143	quiescent	–
z2h600	6.7e11	5.9e4	9	2.9e5	143	violent	–
z2h650	4.0e11	5.9e4	9	2.9e5	143	normal	–
z2h830	5.4e11	5.9e4	9	2.9e5	143	normal	–
z5m09	7.6e8	16.8	0.14	81.9	5.6	quiescent	ultra-high resolution
z5m10	1.3e10	131.6	0.4	655.6	7	normal	ultra-high resolution
z5m10mr	1.4e10	1.1e3	1.9	5.2e3	14	normal	–
z5m11	5.6e10	2.1e3	4.2	1.0e4	14	normal	–

Parameters describing the initial conditions for our simulations (units are physical):

(1) Name: Simulation designation.

(2) M_{halo} : Approximate mass of the main halo at $z = 0$ (**mxx** series), $z = 2$ (**z2hxxx** series), or $z = 6$ (**z5mxx** series).

(3) m_b : Initial baryonic (gas and star) particle mass in the high-resolution region.

(4) ϵ_b : Minimum baryonic force softening (minimum SPH smoothing lengths are comparable or smaller. Force softening is adaptive (mass resolution is fixed).

(5) m_{dm} : Dark matter particle mass in the high-resolution region.

(6) ϵ_{dm} : Minimum dark matter force softening (fixed in physical units at all redshifts).

presented in Hopkins et al. (2014). The simulations **m09** and **m10** are isolated dwarfs, constructed using the method from Oñorbe et al. (2014). Simulations **m10v**, **m11**, **m11v**, **m12q**, **m12i**, and **m13** are chosen to match the initial conditions from the AGORA project (Kim et al. 2014), which will enable future comparisons with a wide range of simulation codes and physics implementations. Simulation **m12v** is based on the initial conditions studied in Kereš & Hernquist (2009) and Faucher-Giguère & Kereš (2011). The simulations with a label ‘v’ have relatively violent merger histories at $z < 2$, while the rest have more typical merger histories. The resolution of these simulations is chosen to scale with the mass of the system to ensure we are able to resolve the giant molecular clouds (GMCs). We also include a separate set of simulations run to $z = 2$ (**z2hxxx** series), which are presented in (Faucher-Giguère et al. 2015). Their main halos are chosen to host Lyman break galaxies (LBG) and cover halo masses $1.9 \times 10^{11} - 1.2 \times 10^{12} M_{\odot}$ at $z = 2$. Finally, we include another series of simulations only run to $z \sim 6$, but with extremely high resolutions (**z5mxx** series). These simulations are presented in Ma et al. (2015). Their main halos cover halo masses from $7.7 \times 10^8 - 5.6 \times 10^{10} M_{\odot}$ at $z = 6$ and these galaxies are believed to contribute most to the cosmic reionization (e.g. Kuhlen & Faucher-Giguère 2012; Robertson et al. 2013). The initial conditions of all the simulations are summarized in Table 1.

In our simulations, gas follows an ionized+atomic+molecular cooling curve from $10 - 10^{10}$ K, including metallicity-dependent fine-structure and molecular cooling at low temperatures and high-temperature metal-line cooling followed species-by-species for 11 separately tracked species (H, He, C, N, O, Ne, Mg, Si, S, Ca,

and Fe; see Wiersma et al. 2009a). At each timestep, the ionization states and cooling rates are determined from a compilation of CLOUDY runs, including a uniform but redshift-dependent photo-ionizing background tabulated in Faucher-Giguère et al. (2009), and photo-ionizing and photo-electric heating from local sources. Gas self-shielding is accounted for with a local Jeans-length approximation, which is consistent with the radiative transfer calculations in Faucher-Giguère et al. (2010).

Star formation is allowed only in dense, molecular, and self-gravitating regions with hydrogen number density above some threshold $n_{\text{th}} = 10 - 100 \text{ cm}^{-3}$ (Hopkins et al. 2013b). Stars form at 100% efficiency per free-fall time when the gas meets these criteria. The self-gravity criterion is physically required to obtain the correct spatial star formation distribution in galaxies (Hopkins et al. 2013b; Padoan & Nordlund 2011), but the galaxy-averaged star formation efficiency is regulated by feedback at much lower values ($\sim 1\%$ per dynamical time, e.g. Faucher-Giguère et al. 2013). We stress that changing these parameters in a reasonable range only yields small and random variations to the global star formation history, as long as feedback is active (see Hopkins et al. 2011, 2012).

Once a star forms, it inherits the metallicity of each tracked species from its parent gas particle. Every star particle is treated as a single stellar population with known mass, age, and metallicity. Then all the feedback quantities, including ionizing photon budgets, luminosities, stellar spectra, supernovae (SNe) rates, mechanical luminosities of stellar winds, metal yields, etc., are directly tabulated from the stellar population models in STARBURST99 (Leitherer et al. 1999), assuming a Kroupa (2002) initial mass function

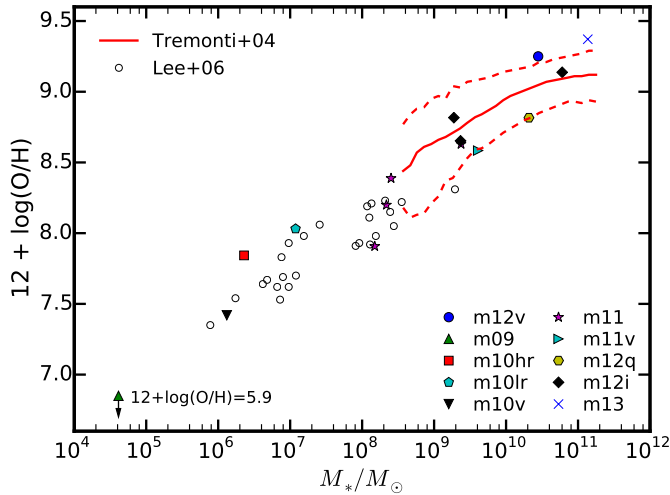


Figure 1. Stellar mass–gas-phase oxygen abundance relation at $z = 0$. The red solid and dashed curves represent the median and 2σ dispersion of the SDSS MZR at $z \sim 0.1$ (Tremonti et al. 2004). The open circles denote the data of the dwarf galaxy sample from Lee et al. (2006). Our simulations are in good agreement with observations from 10^6 – $10^{11} M_\odot$.

(IMF) from 0.1 – $100 M_\odot^2$. We account for several different stellar feedback mechanisms, including: (1) local and long-range momentum flux from radiative pressure; (2) energy, momentum, mass and metal injection from SNe and stellar winds; and (3) photo-ionization and photo-electric heating. We follow Wiersma et al. (2009b) and include the metal yields from Type-II SNe, Type-I SNe, and stellar winds. We note that the Type-II SNe yield table from Woosley & Weaver (1995) adopted in our simulations produce Mg roughly ~ 0.4 dex below the typical values in modern models (e.g. Nomoto et al. 2006). This will have little effect on galaxy properties in our simulations, as Mg is not an important coolant. Nevertheless, we will add 0.4 dex to the Mg abundance to correct this in the analysis below.

All simulations adopt a standard flat Λ CDM cosmology with cosmological parameters consistent with $H_0 = 70.2 \text{ km s}^{-1} \text{ Mpc}^{-1}$, $\Omega_\Lambda = 0.728$, $\Omega_m = 1 - \Omega_\Lambda = 0.272$, $\Omega_b = 0.0455$, $\sigma_8 = 0.807$ and $n = 0.961$ (e.g. Hinshaw et al. 2013; Planck Collaboration et al. 2013).

2.2 Halo Identification, Stellar Mass and Metallicity

We use the Amiga Halo Finder (AHF; Gill et al. 2004; Knollmann & Knebe 2009) to identify galactic halos and galaxies in our simulations. The AHF code uses the adaptive mesh refinement method and identifies halos and subhalos as groups of *bound* particles (dark matter, gas, and stars). In this work, we only consider those “well-resolved” halos that include more than 10^5 bound particles, have at most 10% of their mass contaminated by low-resolution particles, and contain at least 100 gas and 100 star particles, respectively. These criteria are somewhat arbitrary; but varying these numbers within a reasonable range will have little effect on our conclusions.

² In principle, the “IMF-averaged” approximation does not hold for the ultra-high resolution simulations in the **z5mxx** series, where the mass of a star particle is only 10 – $100 M_\odot$. Nevertheless, we confirmed that these simulations predict similar global galaxy properties to those of much poorer resolutions (see Ma et al. 2015).

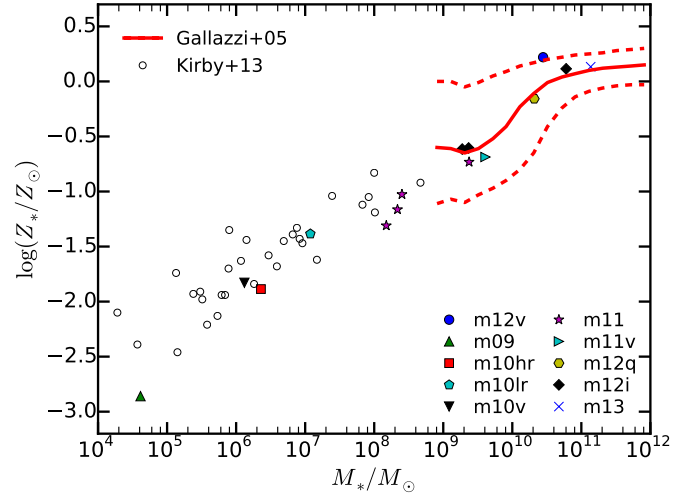


Figure 2. Stellar mass–stellar metallicity relation at $z = 0$. The red solid and dashed curves are the median and 1σ dispersion of the SDSS MZR in the local universe (Gallazzi et al. 2005). The open circles represent the values of $[\text{Fe}/\text{H}]$ of individual dwarfs from Kirby et al. (2013). Again, the agreement is good from 10^4 – $10^{11} M_\odot$.

If none of the halos meets these criteria in a snapshot (this happens in some snapshots at high redshifts ($z \sim 6$), where the galaxy progenitors are too small to contain so many particles), we will take the most massive halo in the high-resolution region in our analysis. We do not include subhalos/satellite galaxies in this work. The centre of a halo is located at the centre of mass of the finest refinement level. We adopt the virial overdensity from Bryan & Norman (1998), which evolves with cosmic time.

We only consider the main galaxy in each halo. To remove the contamination of satellite galaxies, we exclude any gas/star particle that is bound to a subhalo in the analysis below. We measure the galaxy stellar mass (M_*) by summing over the mass of all star particles that belong to the main galaxy. Then we define its stellar metallicity (as well as the abundance of each tracked species) as mass-averaged metallicity (abundance) of all star particles. To separate halo gas and the ISM, we apply a simple temperature criteria and select all gas particles below 10^4 K as the ISM. In our simulations, this is equivalent to selecting gas above some density threshold of a few 0.1 cm^{-3} (we explicitly check the gas distribution in the density–temperature plane), which is comparable to the mean gas density within a few stellar effective radii. It naturally picks warm ionized and cold neutral gas. We define the gas-phase metallicity as the mass-weighted metallicity of all gas particles that belong to the ISM (we compare and discuss three different definitions of gas-phase metallicity in Appendix A)³.

In this work, we use Z_{gas} and Z_* to refer to the mass fraction of all heavy elements in gas and stars, respectively. In Section 3, we will primarily use oxygen abundance $12 + \log(\text{O}/\text{H})$ to present gas-phase metallicities, which is defined in terms of number ratio of oxygen to hydrogen atoms, in order to directly compare with ob-

³ In many cosmological simulations with “sub-grid” models, gas-phase metallicity is usually defined as star-formation-rate-averaged metallicity. However, our simulations explicitly resolve multi-phase ISM structures and include realistic models of star formation and feedback. Individual gas particles are very sensitive to local feedback processes. For these reasons, we do not apply SF-averaged gas-phase metallicity to our simulations.

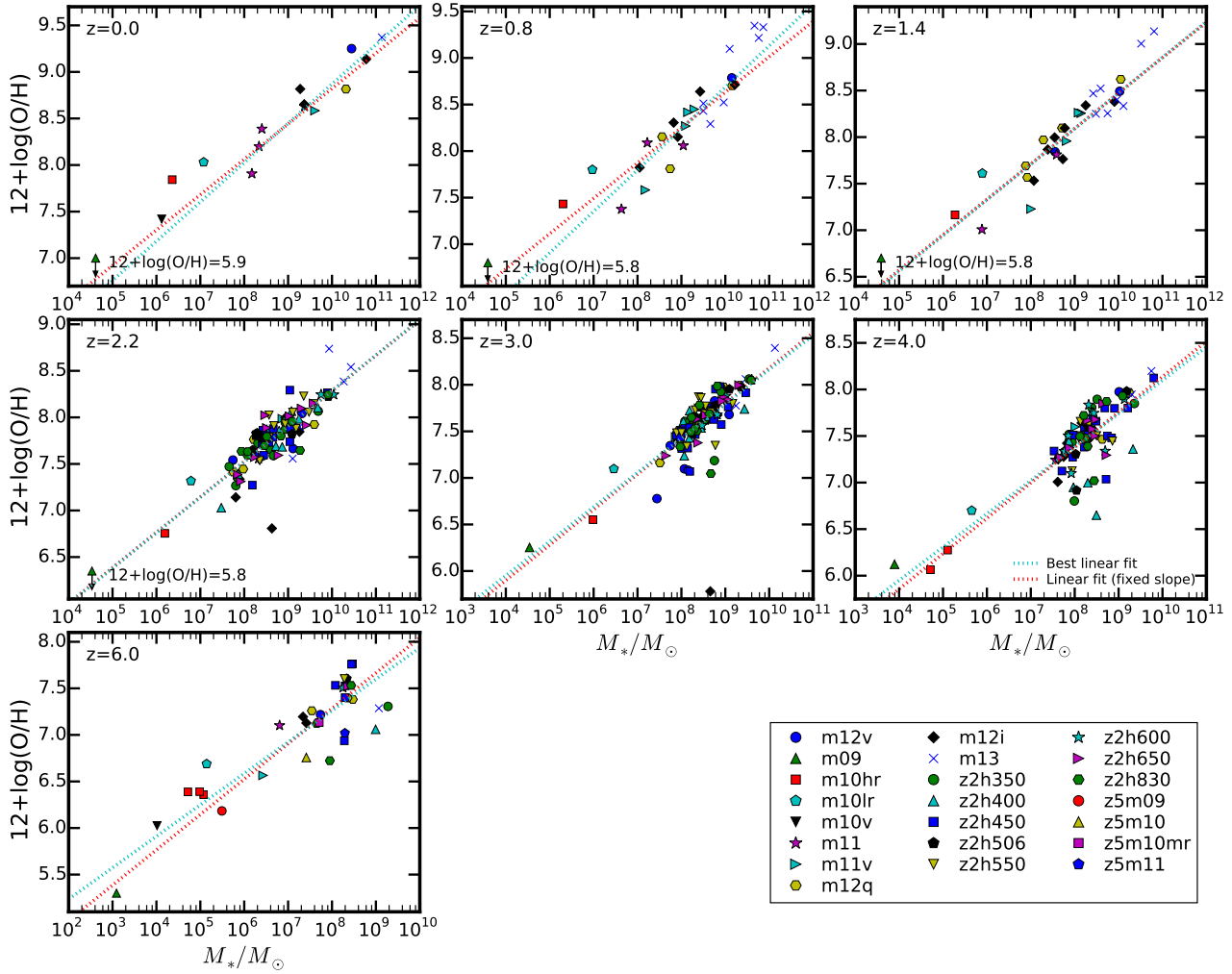


Figure 3. Stellar mass–gas-phase metallicity relation at all redshifts. Cyan dotted lines show the best linear fit $\log(Z_{\text{gas}}/Z_{\odot}) = 12 + \log(\text{O}/\text{H}) - 9.0 = \gamma_g [\log(M_*/M_{\odot}) - 10] + Z_{g,10}$. The red dotted lines show the best fit for a fixed slope $\gamma_g = 0.35$. Note that a constant slope provides a very good fit, where the zero point evolves by ~ 1 dex from $z = 6$ – 0 .

servations. For stellar metallicity, we will primarily use Z_* in the rest of this work. In the literature, gas-phase metallicity and stellar metallicity are also sometimes referred as Z_{gas} and iron abundance $[\text{Fe}/\text{H}]$ (in solar units), respectively. For these reasons, we provide the conversion between these quantities for our simulated galaxies. We will show the calibration in Appendix B but directly give the results here. For a solar metallicity of 0.02 and a solar iron abundance 0.00173 (both in mass fraction), we obtain $12 + \log(\text{O}/\text{H}) = \log(Z_{\text{gas}}/Z_{\odot}) + 9.00$ and $[\text{Fe}/\text{H}] = \log(Z_*/Z_{\odot}) - 0.20$. We emphasize that these relations may suffer from systematic uncertainties that originate from: (1) Type-II and Type-I SNe rates, (2) metal yields of tracked species from different channels, and (3) the solar abundances we adopt in our simulations. However, the shape and evolution of the MZR should be robust to these uncertainties.

3 THE MASS–METALLICITY RELATION

In this section, we present both the gas-phase and stellar MZR from $z = 0$ – 6 and compare our results with observations and other simulations. We will further explore the most important factors that shape the MZR and drive its evolution in the Section 4.

3.1 MZR at $z = 0$

We begin by showing the gas-phase MZR at $z = 0$. In Figure 1, we present the stellar mass–gas-phase oxygen abundance relation for our **mxx** series simulations at $z = 0$. For comparison, we also present the median and 2σ dispersion of the SDSS MZR from Tremonti et al. (2004, red solid and dashed lines) and the data of individual local dwarf galaxies compiled in Lee et al. (2006, open circles) in Figure 1. We remind the reader that these observed gas-phase oxygen abundances are derived from the relative strength of strong nebulae emission lines produced by photo-ionization from young massive stars, so that the observed gas-phase MZR only holds for star-forming galaxies. Also, we emphasize that the overall shape of gas-phase MZR strongly depends on which empirical calibration it uses and the normalization of this relation differs by at most 0.7 dex between different calibrations (Kewley & Ellison 2008, see also Figure 6). For these reasons, we do not apply any correction to these observed data but keep them in their original forms.

In general, our simulations agree reasonably well with observations across stellar mass from $M_* = 10^6$ – $10^{11} M_{\odot}$. The gas-phase metallicity keeps increasing with stellar mass until $M_* \sim 10^{11} M_{\odot}$.

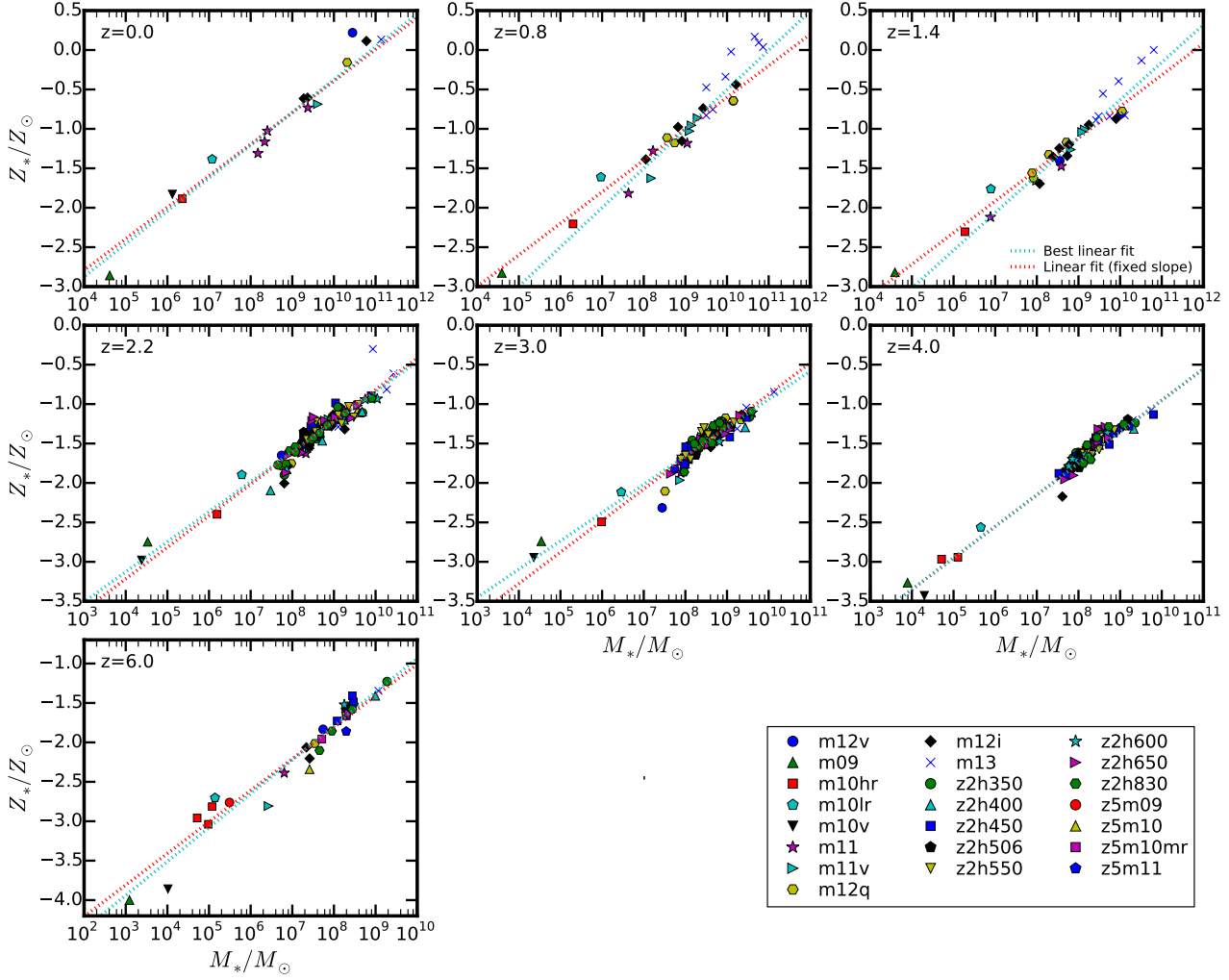


Figure 4. Stellar mass–stellar metallicity relation at all redshifts. Cyan dotted lines show the best linear fit at each redshift $\log(Z_*/Z_\odot) = [\text{Fe}/\text{H}] + 0.2 = \gamma_* [\log(M_*/M_\odot) - 10] + Z_{*,10}$. The red dotted lines show the best fit for a fixed slope $\gamma_* = 0.40$. Again, the slope is approximately constant, while the normalization evolves by ~ 1 dex.

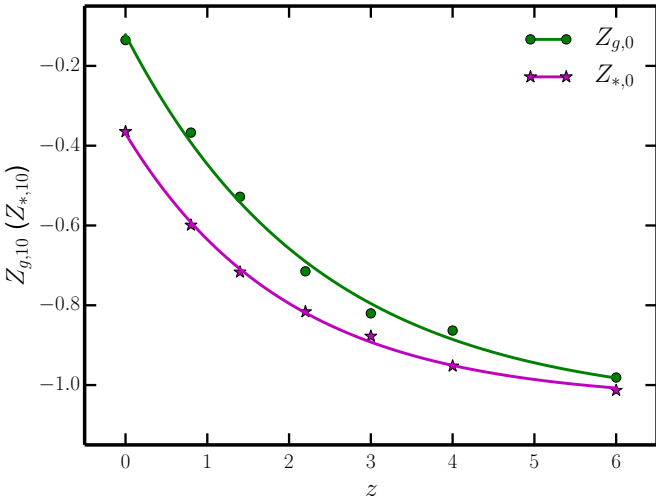


Figure 5. The zero points of gas-phase and stellar MZR, $Z_{g,10}$ and $Z_{*,10}$ as a function of redshift. The solid lines are the best fit of exponential functions $Z_{g,10} = 0.93 \exp(-0.43z) - 1.05$ and $Z_{*,10} = 0.67 \exp(-0.50z) - 1.04$.

and becomes flat at higher masses. Our simulations predict slightly higher metallicities than the observed relation from Tremonti et al. (2004) at the high-mass end. As it has been shown in Hopkins et al. (2014), the most massive galaxies in our simulations ($M_* \gtrsim 10^{11} M_\odot$) have a “quenching problem” at low redshift, i.e., their star formation rates are much higher than observationally inferred values at $z < 1$. Consequently, these galaxies might be over-enriched at low redshift. If so, this suggests that additional physics, such as AGN feedback, is probably required to fully understand the chemical evolution in massive galaxies, at least in the sense of quenching star formation. Alternatively, it has also been proposed that the observed MZR could continue to rise at the high-metallicity end when using new metallicity diagnostics that account for non-equilibrium electron energy distributions (see e.g. Dopita et al. 2013; Nicholls et al. 2013). We emphasize that the “flatness” of MZR at the high-mass end behaves very differently when applying different empirical calibrations (e.g. Kewley & Ellison 2008). Therefore, we do not further quantitatively discuss the discrepancy between our simulations and observations at the massive-end of MZR.

Most of our simulated galaxies are still forming stars (at least very weakly) at $z = 0$, except for **m09**. The **m09** is a low-mass

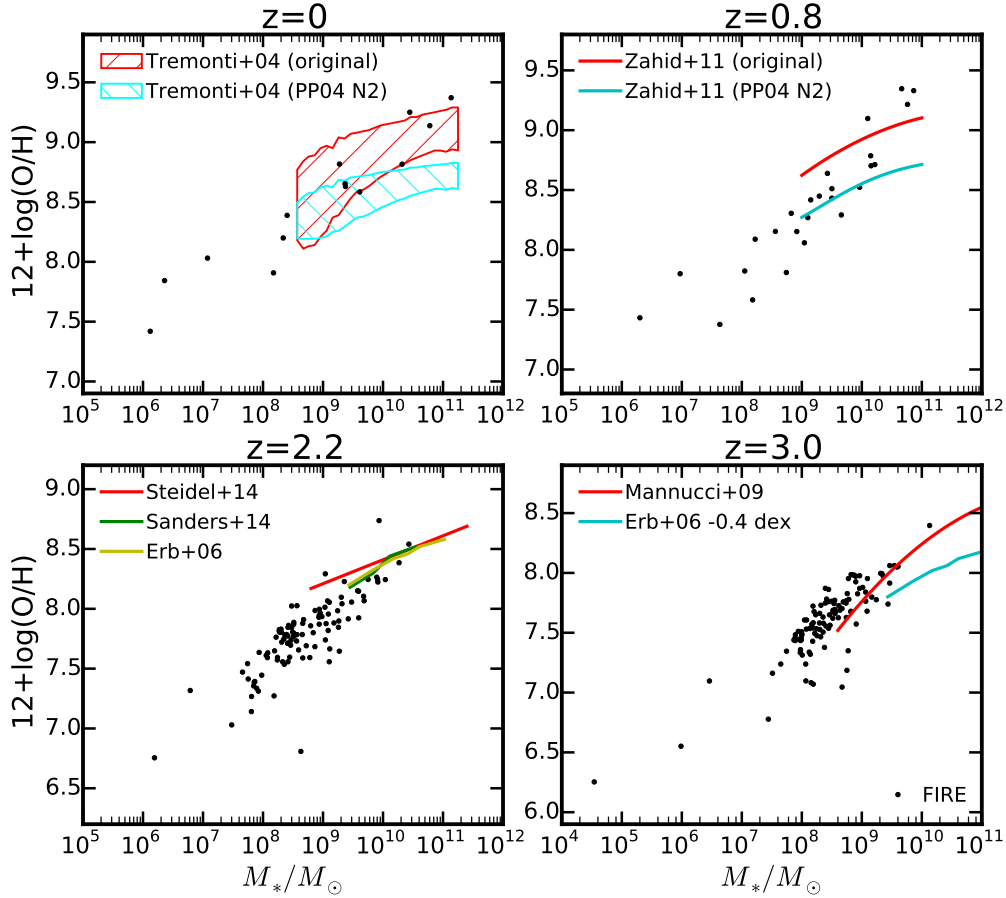


Figure 6. Stellar mass–gas-phase oxygen abundance relations at $z = 0, 0.8, 2.2$, and 3.0 , as compared with a number of observations at these redshifts. In the upper panels, we show both the original relations (red lines) from Tremonti et al. (2004, $z \sim 0$) and Zahid et al. (2011, $z \sim 0.8$) and the relations converted to PP04 N2 calibration (cyan lines) following Kewley & Ellison (2008). In the lower left panel, we show the observed MZR at $z \sim 2.3$ from Steidel et al. (2014, the red line), Sanders et al. (2014, the green line), and Erb et al. (2006, the yellow line). In the lower right panel, we show the best fitting from Mannucci et al. (2009, $z \sim 3.1$). We also shift the Erb et al. (2006) data downward by 0.4 dex for a comparison as motivated by Figure 5 in Mannucci et al. (2009). Given the systematic uncertainties in these observations, our simulations agree reasonably well with the observed evolution.

isolated dwarf galaxy (comparable to the ultra faint dwarfs around the Milky Way), in which star formation has been shut down since $z = 3$ by cosmic reionization (Oñorbe et al. 2015). At $z = 0$, this galaxy has lost almost all metals it produced (see Section 4). Although its gas-phase metallicity is an order of magnitude lower than the extrapolation of the observed MZR down to $M_* = 10^4 M_\odot$, it is not contradictory to observations in the sense that the gas-phase metallicity of such galaxies cannot be measured due to lack of strong nebular emission lines.

In Figure 2, we show the stellar mass–stellar metallicity relationship at $z = 0$ and compare our simulations with the SDSS sample from Gallazzi et al. (2005, red solid and dashed curves) and the dwarf galaxies from Kirby et al. (2013, open circles). Note that the stellar metallicities from Gallazzi et al. (2005) are measured from absorption features of galaxy-integrated spectra (mostly Mg and Fe lines), while the metallicities from Kirby et al. (2013) are derived from Fe abundances of individual stars. The conversion between different methods and their systematic uncertainties are complex and beyond the scope of this paper. For our purposes, we avoid any

correction to these observations but present them in their original values⁴.

Our simulations match these observations quite well over the whole mass coverage from $M_* = 10^4$ – $10^{11} M_\odot$. The simulated sample shows a flatness of stellar MZR above $M_* > 10^{11} M_\odot$ at $z = 0$, consistent with the observed SDSS MZR from Gallazzi et al. (2005). This is the consequence of the fact that the growth of the more massive galaxies in our simulations is dominated by mergers and accretion of low-mass metal-poor satellites rather than *in situ* star formation at low redshifts. Therefore, the average stellar metallicities do not strongly evolve despite the fact that the stellar masses may grow considerably at low redshifts (see also Choi et al. 2014). The gas-phase MZR does not flatten above $M_* > 10^{11} M_\odot$, because gas-phase metallicities still increase considerably due to metal enrichment from Type-Ia SNe and AGB winds even without new star formation.

⁴ In Figure 2, we plot the values of $[\text{Fe}/\text{H}]$ from Kirby et al. (2013), avoiding the complicated conversion between $[\text{Fe}/\text{H}]$ and Z_*/Z_\odot for the observed sample.

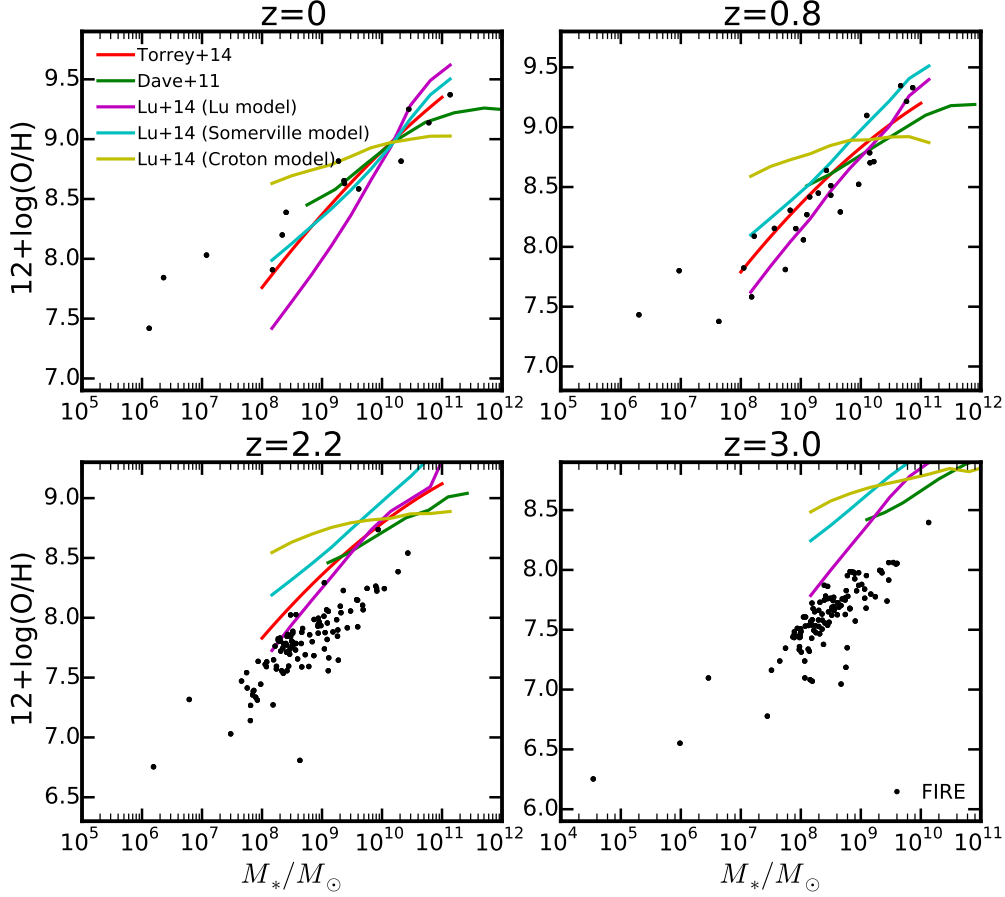


Figure 7. Stellar mass–gas-phase oxygen abundance relation at $z = 0, 0.8, 2.2$, and 3.0 , as compared with other numerical simulations and semi-analytic models. We renormalize other works to $12 + \log(\text{O}/\text{H}) = 8.9$ at $M_* = 10^{10} M_\odot$ at $z = 0$ with respect to our simulations. Red and green lines show the results from cosmological simulations presented in Torrey et al. (2014) and Davé et al. (2011b), respectively, which used popular “sub-grid” models for galactic winds. Magenta, cyan, and yellow lines show the predictions of three semi-analytic models from Lu et al. (2014, the Lu model, the Somerville model, and the Croton model, respectively). All of these models reproduce the correct $z = 0$ stellar mass function, but none of them correctly reproduces the slope or the redshift evolution of the MZR.

3.2 Evolution of the MZR

Figure 3 and 4 show the gas-phase and stellar MZR, respectively, from $z = 0$ –6. We note that for $z \geq 2$ and $z \geq 6$, we include the **z2hxxx** and **z5mxx** simulations in our analysis. The stellar MZR is tighter than the gas-phase MZR, i.e., the gas-phase MZR has larger scatter than stellar MZR at fixed stellar mass. This is because in our simulations, especially at high redshifts, star formation is dominated by multiple bursts, which drives bursts of gas outflows (Muratov et al. 2015). As a consequence, instantaneous gas-phase metallicities may have considerable time fluctuations associated with gas inflows, outflows, and mergers. This effect is larger at high redshifts when the galaxy progenitors are of much lower masses and galaxy mergers are more common, resulting in some outliers that deviate from the main MZR at high redshifts. Despite the short-time-scale fluctuations, both the gas-phase and stellar metallicities increase with time on cosmological time-scales. At all times, gas-phase metallicities are higher than stellar metallicities, since gas-phase metallicities represent the current state of metal enrichment in the galaxies, while stellar metallicities reflect the average galactic metallicities across the whole time. Both metallicities should converge at high redshifts.

To illustrate this quantitatively, we fit the gas-phase and stel-

lar MZR at different redshifts for our simulated galaxies with simple linear functions $\log(Z_{\text{gas}}/Z_\odot) = 12 + \log(\text{O}/\text{H}) - 9.0 = \gamma_g [\log(M_*/M_\odot) - 10] + Z_{g,10}$ and $\log(Z_*/Z_\odot) = [\text{Fe}/\text{H}] + 0.2 = \gamma_* [\log(M_*/M_\odot) - 10] + Z_{*,10}$, where γ_g and γ_* are the slopes and $Z_{g,10}$ and $Z_{*,10}$ represent the typical gas-phase metallicity and stellar metallicity at $M_* = 10^{10} M_\odot$. Although simple linear function do not capture the flatness of stellar metallicity above $M_* \sim 10^{11} M_\odot$ at $z < 1$, it is sufficient for our purposes here. We use least-squares fitting to obtain the best fit (the cyan dotted lines in Figure 3 and 4). In principle, both the slopes and zero points should be functions of redshift. Nevertheless, the MZR at different redshifts have very similar slopes. For simplicity, we pick the mean slope of each relation and redo the linear fit using fixed slopes. We choose $\gamma_g = 0.35$ and $\gamma_* = 0.40$ (red dotted lines in Figure 3 and 4) and confirm that both the best linear fit and the fixed-slope fit describe the simulations reasonably well. We then attribute the evolution of MZR to the evolution of $Z_{g,10}$ and $Z_{*,10}$ with redshift, which we show in Figure 5. We fit these parameters as a function of redshift by an exponential function $F(z) = A \exp(-Bz) + C$. The best fit gives $Z_{g,10} = 0.93 \exp(-0.43z) - 1.05$ and $Z_{*,10} = 0.67 \exp(-0.50z) - 1.04$, respectively (the green and magenta lines in Figure 5). These give the gas-phase and stellar MZR from $z = 0$ –6 as $\log(Z_{\text{gas}}/Z_\odot) = 12 + \log(\text{O}/\text{H}) - 9.0 = 0.35 [\log(M_*/M_\odot) -$

$10] + 0.93 \exp(-0.43z) - 1.05$ and $\log(Z_*/Z_\odot) = [\text{Fe}/\text{H}] + 0.2 = 0.40 [\log(M_*/M_\odot) - 10] + 0.67 \exp(-0.50z) - 1.04$, respectively.

In general, the fitting functions above represent the gas-phase and stellar MZR fairly well for our simulated galaxies, except for the flattening of the stellar MZR above $M_* \sim 10^{11} M_\odot$ at $z = 0$. We emphasize that these results have systematic uncertainties from Type-II and Type-Ia SNe rates, the solar abundance, and the metal yield tables we implement in our simulations. When using these fitting functions, one should notice the uncertainties and make adjustments accordingly.

3.3 Comparison with Observations and Other Simulations

In Figure 6, we compare the gas-phase MZR between our simulations and a number of observations at multiple redshifts. We show the observed MZR at $z \sim 0$ (Tremonti et al. 2004), $z \sim 0.8$ (Zahid et al. 2011), $z \sim 2.2$ (Steidel et al. 2014; Sanders et al. 2014; Erb et al. 2006), and $z \sim 3.1$ (Mannucci et al. 2009). We recall that these observed relations are originally obtained using different calibrations and the systematic uncertainty between different metallicity diagnostics could be up to 0.7 dex (Kewley & Ellison 2008). To illustrate this point, we also convert all the observed relation to the N2 calibration from Pettini & Pagel (2004, PP04 hereafter) unless their original data are already presented using this calibration. For Tremonti et al. (2004) and Zahid et al. (2011), we do the conversion following the formula from Kewley & Ellison (2008, Table 3 therein). In either case, we present both their original relations and the converted relations using PP04 N2 calibration in Figure 6. At $z \sim 2.2$, the observed relations are already presented in PP04 N2 calibration (e.g. Steidel et al. 2014; Sanders et al. 2014; Erb et al. 2006). Mannucci et al. (2009) adopted a very different metallicity calibration, which is established using $z \gtrsim 3$ galaxy samples only. Figure 5 in Mannucci et al. (2009) suggests that the MZR evolves by ~ 0.4 dex from $z \sim 3.1$ to $z \sim 2.2$. Motivated by their results, we also move the $z \sim 2.2$ MZR from Erb et al. (2006) downward by 0.4 dex for a comparison (lower right panel in Figure 6).

In general, our simulations agree reasonably well with these observations in a broad range of stellar mass at $z = 0$ –3, especially when the observed relations are in their original forms. We emphasize that the empirical calibrations developed from the local universe are not necessarily valid for high-redshift galaxies (e.g. Steidel et al. 2014). Therefore, any discrepancies here are well within systematic uncertainties.

In Figure 7, we also compare the MZR from our simulations with other cosmological simulations and semi-analytic models. We compare our results with two other simulations, Torrey et al. (2014, red lines) and Davé et al. (2011b, green lines), and three semi-analytic models from Lu et al. (2014, the Lu model, magenta; the Somerville model, cyan; the Croton model, yellow). These models adopt “sub-grid” empirical models of galactic winds and stellar feedback, which couple some fraction of energy and/or momentum from SNe to the gas, and force certain amount of the gas to escape the galaxy. Note that the metal yields and solar abundance used in different works are not exactly the same, we renormalize all the $z = 0$ MZR to $12 + \log(\text{O}/\text{H}) = 8.9$ at $M_* = 10^{10} M_\odot$ for comparison. At $z = 0$, Torrey et al. (2014) and the Lu model show steeper slopes at the low-mass end, due to the low metal retention efficiency in low-mass galaxies, a consequence of invoking strong outflows to suppress star formation in these galaxies⁵. Some models

⁵ This can be simply illustrated using the “leaky box” model (e.g. Schmidt

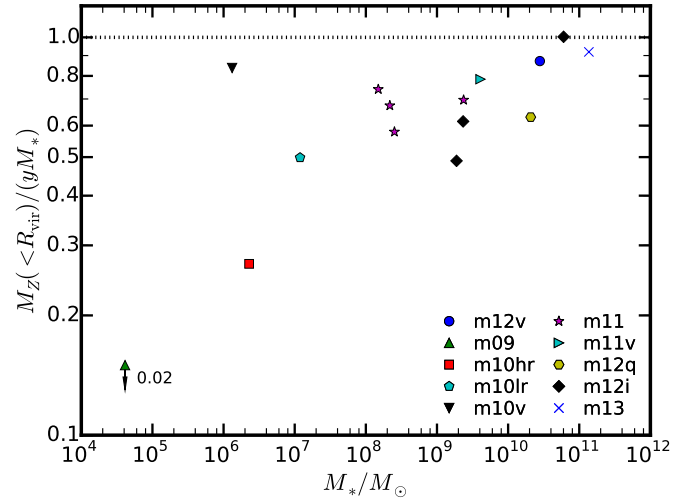


Figure 8. Metal retention fraction for our simulated galaxies at $z = 0$. $M_Z(< R_{\text{vir}})$ is the total amount of metals retained (in both gas and stars) within the virial radius. yM_* (y is the mean effective yield) is the total metal mass produced by stars in the galaxies. The retained fraction in the halo increases with mass and is near unity for all but the least massive system (ultra-faint dwarfs).

predict higher metallicities at the most massive end. Furthermore, these models show significant discrepancies at $z \gtrsim 2$. Our simulations predict much stronger evolution of MZR from $z = 3$ –0 than any other models. Particularly, the Somerville model and the Croton model predict inverse evolution trends – the gas-phase metallicity decreases at lower redshifts at fixed stellar mass – in contrast with observations and other models. We recall that although these models are tuned to match the observed stellar mass function at $z = 0$, they tend to predict systematically higher stellar mass functions than the observed ones for $M_* \lesssim 10^{11} M_\odot$ at $z > 0$ (Somerville & Davé 2014), a consequence of the fact that galaxies in these models form *too* many stars at early time (e.g. Davé et al. 2011a; Sparre et al. 2015; Fiacconi et al. 2015). In Section 4.2, we further explore how the different star formation histories between these models cause the discrepancies in the MZR at high redshifts.

4 DISCUSSION: WHAT DRIVES THE MZR?

In this section, we qualitatively explore what drives the shape and evolution of MZR and try to explain why our simulations differ from other studies.

4.1 Metal retention, gas outflows, and recycling in low-mass galaxies

One major difference between our simulations and some previous work is that our low-mass galaxies are able to retain more metals. To illustrate this, we show in Figure 8 the relation between metal

1963). Assuming the outflow rate is proportional to the star formation rate ($\dot{M}_{\text{out}} = \eta \cdot \text{SFR}$, where η is the mass loading factor), the metallicity is inversely proportional to $1 + \eta$. Low-mass galaxies are more efficient in driving outflows and thus have higher mass loading factors than more massive galaxies. In some simulation with “sub-grid” feedback models and semi-analytic models, low-mass galaxies tend to lose a large fraction of gas (as well as metals) and thus have very low metallicities.

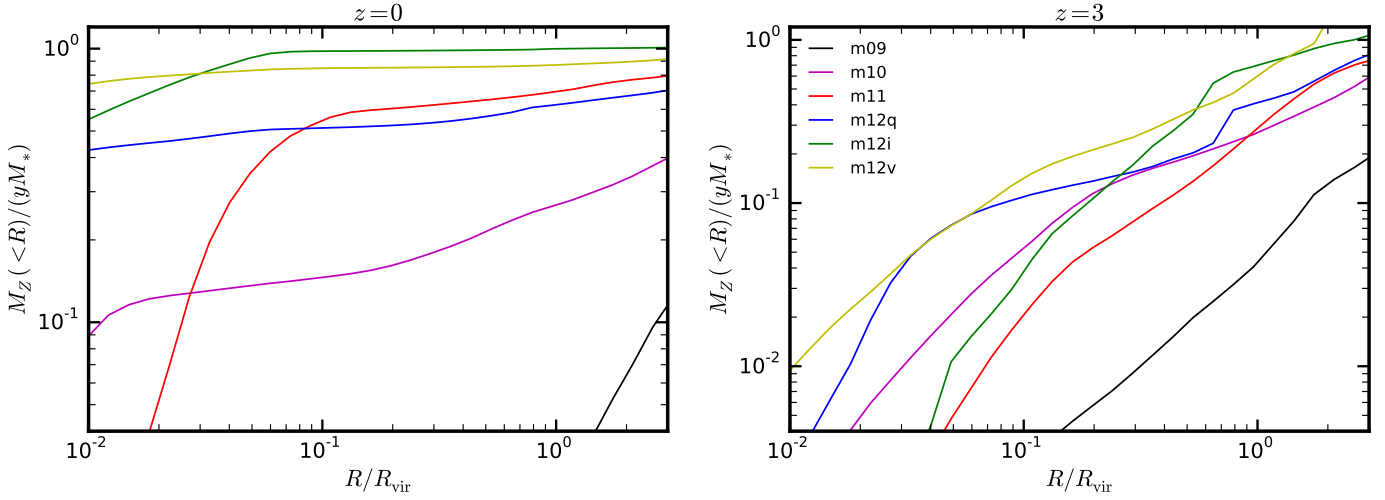


Figure 9. cumulative metal distribution in some galaxies at $z=0$ (left panel) and $z=3$ (right panel), normalized by yM_* (total amount of metal produced by stars in the galaxy). We show **m09** (black), **m10** (magenta), **m11** (red), **m12q** (blue), **m12i** (green), and **m12v** (yellow). In massive galaxies ($M_* > 10^{10} M_\odot$), most metals are retained in stars and the ISM. In ultra-faint dwarfs, reionization and outflows carry most metals away. Intermediate dwarfs reside in between these cases.

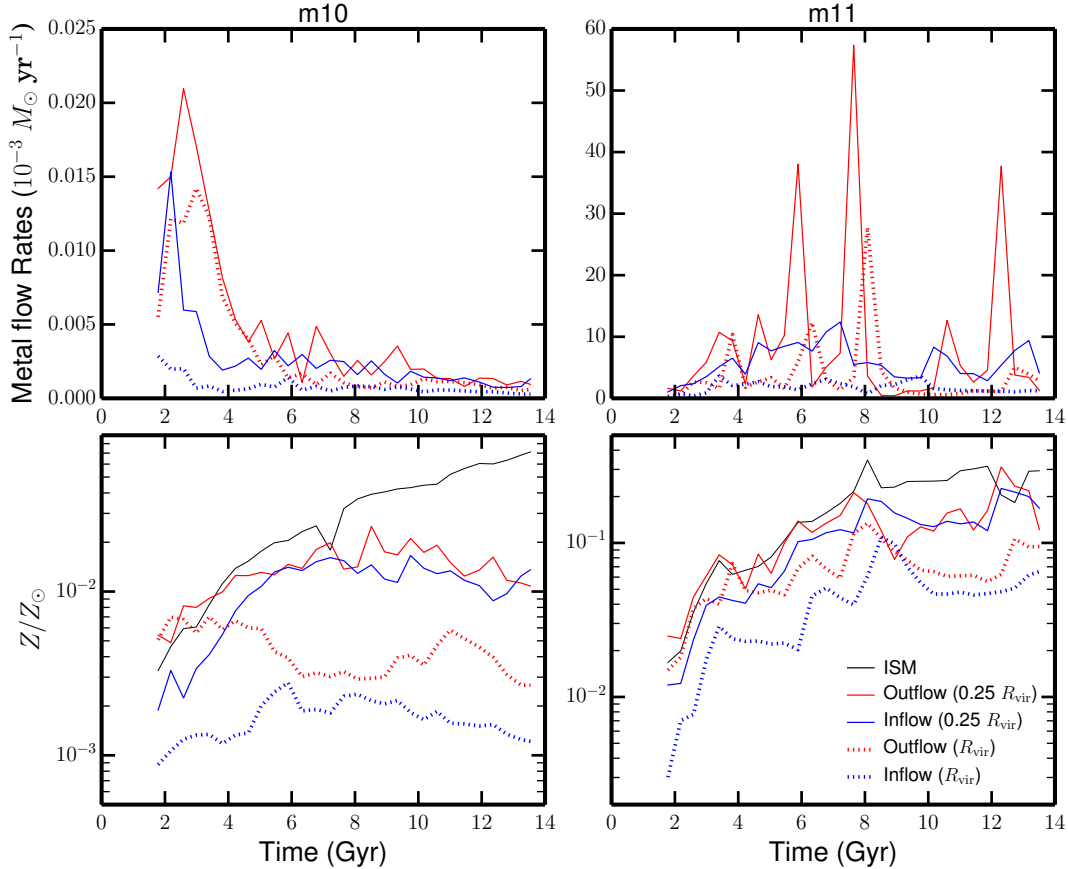


Figure 10. *Upper:* Metal inflow (blue) and outflow rates (red) from $z=0-4$. Solid and dotted lines show the metal inflow/outflow rates measured at $0.25 R_{\text{vir}}$ and R_{vir} , respectively. *Bottom:* Metallicities of inflowing/outflowing gas. The black line shows the metallicity of the ISM. All quantities are averaged over a time-scale of 400 Myr. Metals are efficiently ejected in fountains reaching $0.25 R_{\text{vir}}$, but they do not usually reach R_{vir} – they are either deposited in the halo or recycled efficiently in galactic fountains. Outflowing gas that escapes from the halo at R_{vir} tends to be less enriched than the gas in the ISM.

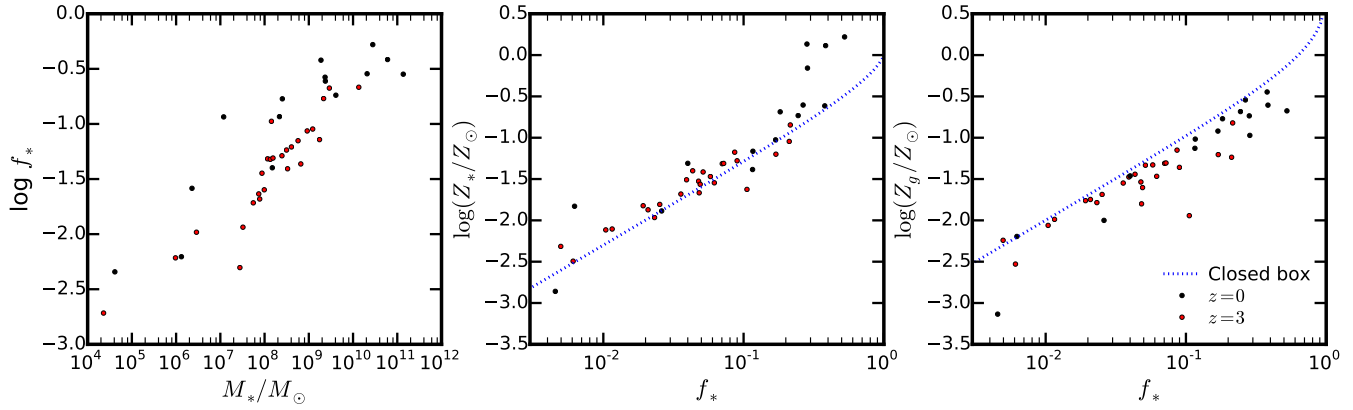


Figure 11. *Left:* Stellar mass fraction $f_* = M_*/(M_{\text{gas}} + M_*)$ as a function of stellar mass. M_{gas} here is the total gas mass in the *halo* (not only in the galaxy). *Middle:* Stellar metallicity Z_* as a function of f_* . For consistency, Z_{gas} here is the average metallicity of *all* gas in the halo (including *both* the ISM *and* the halo gas). Black points and red points show the primary FIRE simulations at $z = 0$ and $z = 3$, respectively. Blue dotted lines show the simple “closed box” model predictions assuming an effective metal yield of $y = 0.02$. The $z = 0$ and $z = 3$ galaxies share the same Z_*-f_* and $Z_{\text{gas}}-f_*$ relations, but the f_*-M_* relation evolves by ~ 0.5 dex from $z = 3-0$. This indicates that the evolution of the MZR is associated with the evolution of f_* (at a fixed stellar mass) at different redshifts. The major offset between our simulations and the predictions of the “closed box” model is largely due to the fact that the metals are not perfectly mixed throughout the halo. Especially in massive galaxies, gas tends to be more metal-enriched in the central star-forming regions than in the outer halo, so stellar metallicities tend to be higher and gas-phase metallicities (including the halo gas) are lower than the predictions of the “closed box” model.

retention fraction and stellar mass at $z = 0$. First, we evaluate the mean effective yield y for every simulation by taking the ratio of total metal mass (in both gas and stars) and total stellar mass in the whole simulation volume. Then for each galaxy, we measure the total metal mass within the virial radius R_{vir} and divide it by yM_* to obtain the metal retention fraction, where M_* is the galaxy stellar mass defined in Section 2.2 and thus yM_* reflects the total amount of metal produced by the stars. As shown in Figure 8, the metal retention fraction in general increases with stellar mass. Galaxies above $M_* > 10^{10.5} M_\odot$ are able to keep almost all metals they have produced. Even at much lower masses ($M_* > 10^6 M_\odot$), galaxies can still retain at least one third or half of their metals within the halo. In contrast, the least massive galaxy in our simulations, **m09**, has lost most of its metals by $z = 0$.

We show the cumulative metal distribution as a function of radius for some of our simulations at $z = 0$ (left panel in Figure 9). The cumulative distributions are normalized by yM_* . For **m11** and more massive galaxies (stellar mass above $10^9 M_\odot$), a large fraction ($\gtrsim 50\%$) of metals are concentrated in the central $0.1 R_{\text{vir}}$ (mostly stars and the ISM). Only a small fraction ($\lesssim 10\%$) of metals are in the halo gas and the rest are lost into the IGM. For **m10** ($M_* = 2 \times 10^6 M_\odot$), metals are more evenly distributed among the galaxy, halo gas, and the IGM. For **m09** ($M_* = 4 \times 10^4 M_\odot$), most of the metals it has produced are in the IGM by $z = 0$. This is consistent with the fact that outflows in low-mass galaxies are stronger (in terms of mass loading) and can propagate more easily to large radii than in more massive systems (Muratov et al. 2015). For a comparison, we also show the cumulative metal distribution at $z = 3$ for the galaxy progenitors (right panel in Figure 9). At high redshift, a significant fraction of metals are still retained in R_{vir} , although metals are more uniformly distributed from the centre to a few virial radii. These galaxies have much lower mass than their low-redshift decedents, and thus they are more efficient in driving gas outflows from star-forming regions throughout the halo.

We further examine the metal inflow/outflow rates and inflow/outflow metallicities. We follow Faucher-Giguère et al. (2011) and Muratov et al. (2015) and define the gas outflow rates, metal

outflow rates, and metallicities of outflow gas as

$$\frac{\partial M}{\partial t} = \sum_i \vec{v} \cdot \frac{\vec{r}}{|r|} M_i / dL, \quad (1)$$

$$\frac{\partial M_{\text{metal}}}{\partial t} = \sum_i \vec{v} \cdot \frac{\vec{r}}{|r|} Z_i M_i / dL, \quad (2)$$

$$Z_{\text{outflow}} = \frac{\partial M_{\text{metal}}}{\partial t} / \frac{\partial M}{\partial t}, \quad (3)$$

where M_i and Z_i are the mass and metallicity of the i^{th} gas particle within the shell of thickness $dL = 0.1 R_{\text{vir}}$ with radial velocity outwards $\vec{v} \cdot \frac{\vec{r}}{|r|} > 0$. The inflow rates and inflow metallicities are defined in the same way but for gas particles with inward radial velocity $\vec{v} \cdot \frac{\vec{r}}{|r|} < 0$. The upper panels in Figure 10 show the metal inflow/outflow rates at $0.25 R_{\text{vir}}$ (blue/red solid lines) and at R_{vir} (blue/red dotted lines) for our **m10** (left) and **m11** (right) simulations. We average the inflow/outflow rates on a time-scale of 400 Myr. In either case, the net metal outflow rates are considerably lower at R_{vir} than at $0.25 R_{\text{vir}}$, indicating that the metals are either deposited in the halo or returned back to the ISM. At high redshifts, metals ejected in outflows can be more easily driven to R_{vir} than at low redshifts. At $0.25 R_{\text{vir}}$, metal inflow rates are comparable to metal outflow rates, suggesting a high efficiency of metal recycling. The lower panels in Figure 10 show the average metallicities of inflows and outflows at both $0.25 R_{\text{vir}}$ and at R_{vir} , as compared to the metallicity of the ISM (black solid lines). The outflow metallicities are much lower at R_{vir} than at $0.25 R_{\text{vir}}$, because outflowing gas sweeps out and mixes with more metal-poor gas in the halo when propagating outwards. Since our simulations explicitly resolve the generation, propagation, mixing, and recycling of gas and metals in galactic fountains on different scales, we produce a low-mass-end slope of the MZR that is closer to observations. A comprehensive analysis on the metallicities of outflows and how they depend on galaxy mass, and the efficiency and time-scale of metal recycling will be presented in future work (Muratov et al., in preparation; Anglés-Alcázar et al., in preparation).

4.2 Stellar mass fraction

Another major difference between our simulations and other theoretical work is we predict much stronger evolution of the MZR from $z = 3-0$ (e.g. the stellar metallicity increases by 0.5 dex at fixed stellar mass, see Figure 5). Observations and some analytic models suggest a fundamental metallicity relation (FMR) between stellar mass, star formation rate, and metallicity that holds for star-forming galaxies both in the local universe and at high redshifts (e.g. Mannucci et al. 2010, 2011; Lilly et al. 2013; Zahid et al. 2014). Motivated by these results, we attempt to qualitatively illustrate what might be the primary factor that drives the evolution of MZR in this section. We start by reviewing the simplest chemical evolution model, i.e., the “closed box” model, which predicts the stellar and gas-phase metallicities as a function of stellar mass fraction, $f_* = M_*/(M_{\text{gas}} + M_*)$ as the following

$$Z_* = y \left[\frac{1-f_*}{f_*} \ln(1-f_*) + 1 \right], \quad (4)$$

$$Z_g = y \ln(1-f_*), \quad (5)$$

where y is the effective metal yield (e.g. Schmidt 1963; Talbot & Arnett 1971; Searle & Sargent 1972). The parameter f_* describes the fraction of baryons that have been turned into stars, and $1 - f_*$ is the “gas fraction”. In Figure 11, we show the relation between stellar and gas-phase metallicities and f_* , respectively (the middle and right panels), for our **mx** series simulations at $z = 0$ and $z = 3$ (black and red points). We emphasize that we account for *both* the halo gas *and* the ISM in the total gas mass when calculating f_* , since halo gas is actively involved in supplying star formation and metal exchange in most cases. For consistency, the gas-phase metallicities shown in Figure 11 is the average metallicity of *all* gas in the halo. For illustrative purpose, we also show the simple predictions from the “closed box” model, assuming an effective metal yield of $y = 0.02$ (blue dotted lines in Figure 11).

The simulated data at $z = 0$ and $z = 3$ overlap with each other in the Z_*-f_* and $Z_{\text{gas}}-f_*$ diagrams. In the left panel of Figure 11, we also show the relation between f_* and M_* for these galaxies at both redshifts. There is a systematic offset (~ 0.5 dex) in the f_*-M_* relation between galaxies at $z = 0$ and 3. Note that in the limit of $f_* \ll 1$, one has $Z_*, Z_{\text{gas}} \propto f_*$. Therefore, the 0.5 dex offset in f_*-M_* relation propagates to the 0.5 dex evolution of the MZR from $z = 3$ to 0. This suggests that the evolution of the MZR is associated with the evolution of f_* (at a fixed stellar mass) within the *halo* at different redshifts, providing a first hint of a universal metallicity relation between stellar mass, gas mass, and metallicities (see also Zahid et al. 2014). In simulations with “sub-grid” feedback models and semi-analytic models, where the $z = 0$ stellar mass functions are tuned to match observations, galaxies tend to form a large fraction of their stars at high redshift and therefore their evolution is weaker at lower redshift (e.g. Somerville & Davé 2014), as opposed to observations and our simulations. In other words, these models produce higher f_* than our simulations at fixed stellar mass at $z > 0$ and an f_*-M_* relation barely evolving from $z = 3-0$. Therefore, galaxies in those models are more metal-enriched at high redshifts and the evolution of the MZR is weaker than our simulations.

Our simulations are qualitatively consistent with the simple “closed box” predictions applied to *halo* quantities⁶. This is not unreasonable because a large fraction (order unity) of metals are

retained within the virial radius at both redshifts (see e.g., Figure 9). However, we emphasize that one should not think our simulated galaxies are closed boxes, because the metals are not perfectly well-mixed in the galactic halo. This explains the major offset between the “closed box” model and our simulations (middle and right panels in Figure 11), especially in the most massive systems where this effect is stronger. Since gas in the centre of the galaxy tends to be more metal-enriched than gas in the outer halo and stars preferentially form in the central region, stellar metallicities tend to be higher and the gas-phase metallicities (including the halo gas) are lower than the predictions of the closed box model (applied to halo quantities). The mixing of metals is very complex and associated with galactic fountains on different scales. Although the “closed box” model gives a natural relation between stellar mass, gas mass, and the metallicities, the parameterization of a universal metallicity relation for galactic quantities (i.e., excluding the halo) is more complicated than the simple model. This is worth further investigation in more detail in future work.

5 CONCLUSION

We use a series of high-resolution cosmological zoom-in simulations spanning halo masses $10^9-10^{13} M_\odot$ and stellar masses $10^4-10^{11} M_\odot$ at $z = 0$ from the FIRE project to study the galaxy mass-metallicity relations at $z = 0-6$. These simulations include explicit models of multi-phase interstellar medium, star formation, and stellar feedback. As has been shown in previous papers, these simulations successfully reproduce many observed galaxy properties, including the stellar mass-halo mass relation, star-forming main sequence, the Kennicutt-Schmidt law, star formation histories, etc., for a wide range of galaxies at many redshifts (Hopkins et al. 2014). These simulations also predict reasonable covering fractions of neutral hydrogen in the halos of $z = 2-3$ LBGs (Faucher-Giguère et al. 2015) and self-consistently generate galactic winds with velocities and mass loading factors broadly consistent with observational requirements (Muratov et al. 2015). These simulations adopt “standard” stellar population models and metal yield tables from Type-I and Type-II supernovae and stellar winds, following species-by-species for 11 separately tracked elements. Our key conclusions include the following.

(i) The simulations predict galaxy mass-metallicity relations that agree reasonably well with a number of observations from $z = 0-3$ for a broad range of stellar masses. Both gas-phase and stellar metallicities evolve monotonically from $z = 0-6$, with higher metal abundance at low redshifts at fixed stellar mass. The best linear fits of the MZR for our simulated galaxies as a function of redshift are $\log(Z_{\text{gas}}/Z_\odot) = 12 + \log(\text{O}/\text{H}) - 9.0 = 0.35 [\log(M_*/M_\odot) - 10] + 0.93 \exp(-0.43z) - 1.05$ and $\log(Z_*/Z_\odot) = [\text{Fe}/\text{H}] + 0.2 = 0.40 [\log(M_*/M_\odot) - 10] + 0.67 \exp(-0.50z) - 1.04$, for gas-phase metallicity and stellar metallicity, respectively. We emphasize that the normalizations may have systematic uncertainties that originate from the SNe rates, yield tables, and solar abundance we adopt, but the evolution of the MZR is robust to these uncertainties.

(ii) The stellar MZR becomes flat around $M_* \sim 10^{11} M_\odot$ since $z = 0$, because the most massive galaxies in our simulations evolve via mergers and accretion of satellites rather than *in situ* star formation at low redshifts. Therefore, the stellar metallicity does not

⁶ We emphasize that in the analog of Figure 11 where we measure f_* using *only* the gas in the galaxy (i.e., excluding the halo gas), all the galaxies are well below the predictions of the closed box model and there is no well-

defined relation, indicating that galaxies themselves are far from closed boxes. This suggests the necessity of accounting for halo gas as reservoirs in galaxy evolution.

increase despite the fact that the stellar mass grows considerably. The gas-phase MZR does not flatten at the high-mass end, because gas keeps being enriched by Type-Ia SNe and AGB winds from old stellar populations.

(iii) The evolution of MZR is associated with the evolution of the gas/stellar mass fraction within the inner halo (not just inside the galaxy effective radius) at different redshifts. This provides a first hint of a universal metallicity relation between stellar mass, gas mass, and metallicities, but its parameterization for galactic quantities (as opposed to for halo quantities, which behave more like a closed box) is much more complicated than simple analytic models. We will investigate this in more detail in future work.

(iv) Galaxies above $10^6 M_\odot$ can retain a large fraction of their metals in the halo even up to $z = 3$. The net metal outflow rates near the virial radius are always lower than those near the galaxy, indicating that the metals either get deposited in the halo or return back to the ISM. The high metal inflow rates and the high metallicity of inflowing gas at $0.25 R_{\text{vir}}$ suggest a high efficiency of metal recycling. On average, the outflows at outer radii are much less metal-enriched than those at the inner radius.

(v) These differential recycling and metal retention effects are not properly accounted for in most semi-analytic and “sub-grid” feedback models that are popular in cosmological simulations. As a result, these simplified models cannot simultaneously reproduce the galaxy mass function and the slope and redshift evolution of the MZR. By explicitly resolving the “missing physics” in these models, we reconcile the long-standing discrepancy, and provide a clear way forward to improve the sub-grid and semi-analytic models.

Nevertheless, our simulations are still limited in sample size. In the near future, we will expand our simulations to include more dwarf galaxies covering halo mass from $M_{\text{halo}} = 10^8 - 10^{11} M_\odot$ and to enlarge our sample at the most massive end to better understand whether the flattening of the MZR is real and what drives the flatness. This may depend critically on AGN feedback. We will provide quantitative analysis on metal outflow rates, outflow metallicities, metal recycling, and their relation with galaxy properties in future work (Muratov et al., in preparation; Anglés-Alcázar et al., in preparation).

ACKNOWLEDGMENTS

We thank Daniel Anglés-Alcázar, Yu Lu, Evan Kirby, Paul Torrey, Andrew Wetzel, and many friends for helpful discussion and useful comments on this paper. The simulations used in this paper were run on XSEDE computational resources (allocations TG-AST120025, TG-AST130039, and TG-AST140023). Support for PFH was provided by the Gordon and Betty Moore Foundation through Grant 776 to the Caltech Moore Center for Theoretical Cosmology and Physics, by the Alfred P. Sloan Foundation through Sloan Research Fellowship BR2014-022, and by NSF through grant AST-1411920. CAFG was supported by a fellowship from the Miller Institute for Basic Research in Science, by NASA through Einstein Postdoctoral Fellowship Award PF3-140106 and grant 10-ATP10-0187, by NSF through grant AST-1412836, and by Northwestern University funds. DK was supported by NSF grant AST-1412153 and UC San Diego funds. EQ was supported by NASA ATP grant 12-APT12-0183, a Simons Investigator award from the Simons Foundation, the David and Lucile Packard Foundation, and the Thomas Alison Schneider Chair in Physics at UC Berkeley.

REFERENCES

- Behroozi, P. S., Wechsler, R. H., & Conroy, C. 2013, *ApJ*, 770, 57
- Belli, S., Jones, T., Ellis, R. S., & Richard, J. 2013, *ApJ*, 772, 141
- Benson, A. J. 2012, *NewA*, 17, 175
- Bertone, S., De Lucia, G., & Thomas, P. A. 2007, *MNRAS*, 379, 1143
- Bryan, G. L., & Norman, M. L. 1998, *ApJ*, 495, 80
- Choi, J., Conroy, C., Moustakas, J., Graves, G. J., Holden, B. P., Brodwin, M., Brown, M. J. I. & van Dokkum, P. G. et al. 2014, *ApJ*, 792, 95
- Croton, D. J. et al. 2006, *MNRAS*, 365, 11
- Davé, R., Oppenheimer, B. D., & Finlator, K. 2011, *MNRAS*, 415, 11
- Davé, R., Finlator, K., & Oppenheimer, B. D. 2011, *MNRAS*, 416, 1354
- Davé, R., Finlator, K., & Oppenheimer, B. D. 2012, *MNRAS*, 421, 98
- Dekel, A., & Silk, J. 1986, *ApJ*, 303, 39
- Dopita, M. A., Sutherland, R. S., Nicholls, D. C., Kewley, L. J., & Vogt, F. P. A. 2013, *ApJS*, 208, 10
- Edmunds, M. G. 1990, *MNRAS*, 246, 678
- Erb, D. K., Shapley, A. E., Pettini, M., Steidel, C. C., Reddy, N. A., & Adelberger, K. L. 2006, *ApJ*, 644, 813
- Faucher-Giguère, C.-A., Lidz, A., Zaldarriaga, M., & Hernquist, L. 2009, *ApJ*, 703, 1416
- Faucher-Giguère, C.-A., Kereš, D., Dijkstra, M., Hernquist, L., & Zaldarriaga, M. 2010, *ApJ*, 725, 633
- Faucher-Giguère, C.-A., & Kereš, D. 2011, *MNRAS*, 412, L118
- Faucher-Giguère, C.-A., Kereš, D., & Ma, C.-P. 2011, *MNRAS*, 417, 2982
- Faucher-Giguère, C.-A., Quataert, E., & Hopkins, P. F. 2013, *MNRAS*, 433, 1970
- Faucher-Giguère, C.-A., Hopkins, P. F., Kereš, D., Muratov, A. L., Quataert, E., & Murray, N. 2015, *MNRAS*, 449, 987
- Fiacconi, D., Feldmann, R., & Mayer, L. 2015, *MNRAS*, 446, 1957
- Finlator, K., & Davé, R. 2008, *MNRAS*, 385, 2181
- Gallazzi, A., Charlot, S., Brinchmann, J., White, S. D. M., & Tremonti, C. A. 2005, *MNRAS*, 362, 41
- Gill, S. P. D., Knebe, A., & Gibson, B. K. 2004, *MNRAS*, 351, 399
- Hahn, O., & Abel, T. 2011, *MNRAS*, 415, 2101
- Henry, A., Martin, C. L., Finlator, K., & Dressler, A. 2013, *ApJ*, 769, 148
- Henry, A. et al. 2013, *ApJ*, 776, L27
- Hinshaw, G. et al. 2013, *ApJS*, 208, 19
- Hopkins, P. F., Quataert, E., & Murray, N. 2011, *MNRAS*, 417, 950
- Hopkins, P. F., Quataert, E., & Murray, N. 2012, *MNRAS*, 421, 3522
- Hopkins, P. F. 2013, *MNRAS*, 428, 2840
- Hopkins, P. F., Cox, T. J., Hernquist, L., Narayanan, D., Hayward, C. C. 2013, *Murray, N. MNRAS*, 430, 1901
- Hopkins, P. F., Narayanan, D., & Murray, N. 2013, *MNRAS*, 432, 2647
- Hopkins, P. F., Keres, D., Onorbe, J., Faucher-Giguère, C.-A., Quataert, E., Murray, N., & Bullock, J. S. 2014, *MNRAS*, 445, 581
- Hopkins, P. F. 2014, *arXiv:1409.7395*
- Kereš, D., Katz, N., Weinberg, D. H., & Davé, R. 2005, *MNRAS*, 363, 2

Kereš, D., & Hernquist, L. 2009, *ApJ*, 700, L1
 Kewley, L. J., & Ellison, S. L. 2008, *ApJ*, 681, 1183
 Kim, J.-h. et al. 2014, *ApJS*, 210, 14
 Kirby, E. N., Cohen, J. G., Guhathakurta, P., Cheng, L., Bullock, J. S., Gallazzi, A. 2013, *ApJ*, 779, 102
 Knollmann, S. R., & Knebe, A. 2009, *ApJS*, 182, 608
 Kroupa, P. 2002, *Science*, 295, 82
 Kuhlen, M., & Faucher-Giguère, C.-A. 2012, *MNRAS*, 423, 862
 Lee, H., Skillman, E. D., Cannon, J. M., Jackson, D. C., Gehrz, R. D., Polomski, E. F., & Woodward, C. E. 2006, *ApJ*, 647, 970
 Leitherer, C., et al. 1999, *ApJS*, 123, 3
 Lilly, S. J., Carollo, C. M., Pipino, A., Renzini, A., & Peng, Y. 2013, *ApJ*, 772, 119
 Lu et al. 2014, *ApJ*, 795, 123
 Ma, X., Kasen, D., Hopkins, P. F., Faucher-Giguère, C.-A., Quataert, E., Kereš, D., Murray, N. 2015, arXiv:1503.07880
 Maiolino, R. et al. 2008, *A&A*, 488, 463
 Mannucci, F., Della Valle, M., & Panagia, N. 2006, *MNRAS*, 370, 773
 Mannucci, F. et al. 2009, *MNRAS*, 398, 1915
 Mannucci, F., Cresci, G., Maiolino, R., Marconi, A., & Gnerucci, A. 2010, *MNRAS*, 408, 2115
 Mannucci, F., Salvaterra, R., & Campisi, M. A. 2011, *MNRAS*, 414, 1263
 Moster, B. P., Naab, T., & White, S. D. M. 2013, *MNRAS*, 428, 3121
 Muratov, A. L., Keres, D., Faucher-Giguere, C.-A., Hopkins, P. F., Quataert, E., Norman, N. 2015, arXiv:1501.03155
 Nicholls, D. C., Dopita, M. A., Sutherland, R. S., Kewley, L. J., & Palay, E. 2013, *ApJS*, 207, 21
 Nomoto, K., Tominaga, N., Umeda, H., Kobayashi, C., & Maeda, K. 2006, *Nuclear Physics A*, 777, 424
 Oñorbe, J., Garrison-Kimmel, S., Maller, A. H., Bullock, J. S., Rocha, M., & Hahn, O. 2014, *MNRAS*, 437, 1894
 Oñorbe, J., Boylan-Kolchin, M., Bullock, J. S., Hopkins, P. F., Kereš, D., Faucher-Giguère, C.-A., Quataert, E., Murray, N., 2015, arXiv:1502.02036
 Oppenheimer, B. D., Davé, R., Kereš, D., et al. 2010, *MNRAS*, 406, 2325
 Padoan, P., & Nordlund, Å. 2011, *ApJ*, 730, 40
 Pettini, M., & Pagel, B. E. J. 2004, *MNRAS*, 348, L59
 Planck Collaboration et al. 2013, arXiv:1303.5076
 Robertson, B. E. et al. 2013, *ApJ*, 768, 71
 Sanders, R. L. et al. 2014, arXiv:1408.2521
 Scannapieco, C. et al. 2012, *MNRAS*, 423, 1726
 Schmidt, M. 1963, *ApJ*, 137, 758
 Searle, L., & Sargent, W. L. W. 1972, *ApJ*, 173, 25
 Somerville, R. S., Hopkins, P. F., Cox, T. J., Robertson, B. E., & Hernquist, L. 2008, *MNRAS*, 391, 481
 Somerville, R. S., & Davé, R. 2014, arXiv:1412.2712
 Sparre, M. et al. 2015, *MNRAS*, 447, 3548
 Springel, V. 2005, *MNRAS*, 364, 1105
 Springel, V., Di Matteo, T., & Hernquist, L. 2005, *ApJ*, 620, L79
 Steidel, C. C. et al. 2014, *ApJ*, 795, 165
 Talbot, R. J., Jr., & Arnett, W. D. 1971, *ApJ*, 170, 409
 Torrey, P., Vogelsberger, M., Genel, S., Sijacki, D., Springel, V., & Hernquist, L. 2014, *MNRAS*, 438, 1985
 Tremonti, C. A. et al. 2004, *ApJ*, 613, 898
 Veilleux, S., Cecil, G., & Bland-Hawthorn, J. 2005, *ARA&A*, 43, 769
 Wiersma, R. P. C., Schaye, J., & Smith, B. D. 2009, *MNRAS*, 393, 99

Wiersma, R. P. C., Schaye, J., Theuns, T., Dalla Vecchia, C., & Tornatore, L. 2009, *MNRAS*, 399, 574
 Woosley, S. E., & Weaver, T. A. 1995, *ApJS*, 101, 181
 Yabe, K., Ohta, K., Iwamuro, F., et al. 2014, *MNRAS*, 437, 3647
 Zahid, H. J., Kewley, L. J., & Bresolin, F. 2011, *ApJ*, 730, 137
 Zahid, H. J., Geller, M. J., Kewley, L. J., Hwang, H. S., Fabricant, D. G., Kurtz, M. J. 2013, *ApJ*, 771, L19
 Zahid, et al., Dima, G. I., Kudritzki, R.-P., Kewley, L. J., Geller, M. J., Hwang, H. S., Silverman, J. D., & Kashino, D. 2014, *ApJ*, 791, 130

APPENDIX A: DIFFERENT DEFINITIONS OF GAS-PHASE METALLICITY

In this work, the gas-phase metallicity is defined as the mass-weighted average metallicity of all gas particles below 10^4 K, which we refer as the ISM gas. In principle, there are many alternative approaches to define gas-phase metallicities. In this section, we discuss three definitions and compare them with each other: (1) the average metallicity of all gas particles below 10^4 K in the galaxy (our default definition), (2) the average metallicity of all gas particles within $0.1 R_{\text{vir}}$, and (3) the average metallicity of all gas particles with temperature between 7,000–15,000 K and density above 0.5 cm^{-3} . In Figure 1, we compare definition (1) and (2) in the left panel and (1) and (3) in the right panel for all galaxies presented in Figure 3.

Definition (1) is designed to automatically select all the warm ionized gas and cold neutral gas (the ISM), definition (2) aims to pick the gas in the star-forming regions, and definition (3) is observationally motivated to select the nebular gas which produce the strong nebular emission lines in star-forming galaxies. In general, these definitions are consistent with each other. Most of the galaxies lie very close to the $y = x$ relation in each panel of Figure 1. However, there are a few outliers in these diagrams. Definition (2) can be problematic in merging systems, where the halo centre may deviate far from the stellar bulk and thus $0.1 R_{\text{vir}}$ does not necessarily probe the star-forming region. Definition (3) is largely affected by abundance variance between gas particles, since there are usually not many gas particles at any single instant that meet the temperature and density criteria. However, a time-averaged version of definition (3) removes most of the outliers. Therefore, we argue that our default definition is more adaptive and flexible than other definitions.

APPENDIX B: METALLICITIES IN DIFFERENT FORMS

In this work, we primarily use $12 + \log(\text{O}/\text{H})$ and Z_* to present gas-phase metallicity and stellar metallicity, respectively. In the literature, gas-phase metallicity and stellar metallicity are sometimes presented in terms of Z_{gas} and $[\text{Fe}/\text{H}]$. Therefore, we also provide the conversion between these different forms of metallicities for comparison. We emphasize these conversions are obtained from our simulations only and there are systematic uncertainties originating from the uncertain relative metal yields between species and solar abundances we adopt.

In Figure 2, we show the relations between $12 + \log(\text{O}/\text{H})$ and $\log(Z_{\text{gas}}/Z_{\odot})$ (left panel) and the relation between $[\text{Fe}/\text{H}]$ and $\log(Z_*/Z_{\odot})$ (right panel), where we adopt a solar metallicity $Z_{\odot} = 0.02$ and a solar iron abundance of 0.00173, both in mass fraction. In both panels, we collect data of all the simulated galaxies at all

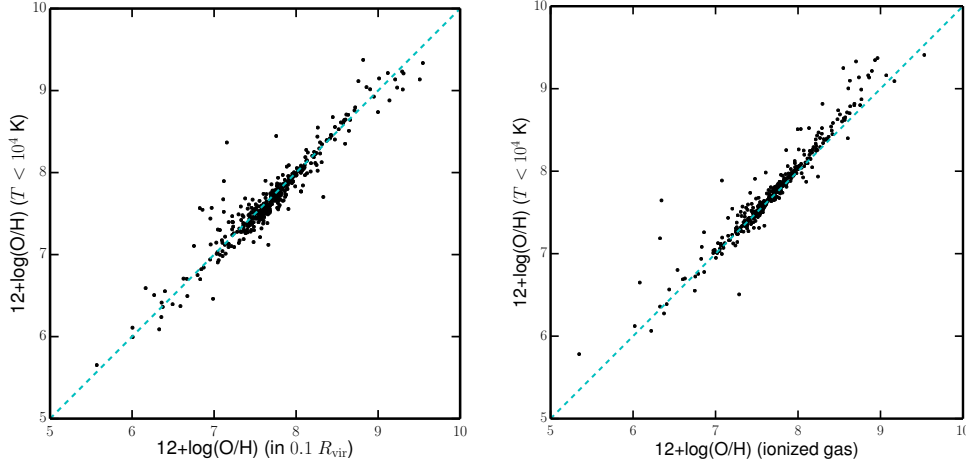


Figure 1. Gas-phase oxygen abundances in different definitions. *Left:* The relation of gas oxygen abundances between definition (1) the average metallicity of all gas particles below 10^4 K and (2) the average metallicity of all gas particles within $0.1 R_{\text{vir}}$. *Right:* The relation of gas oxygen abundances between definition (1) and (3) the average metallicity of all gas particles with temperature between 7,000–15,000 K and density above 0.5 cm^{-3} . The cyan dashed lines show the $y = x$ relation. The black points show all the data presented in Figure 3. Different definitions agree well, and have *no* qualitative effect on any of our conclusions. Most of the “outliers” are caused by transient, stochastic time variability.

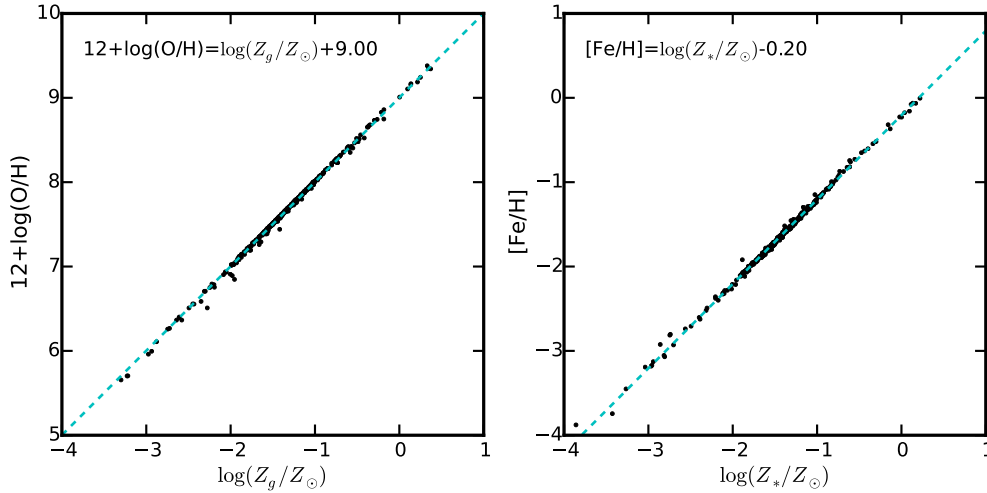


Figure 2. Relations between different forms of metallicities. *Left:* Gas-phase oxygen abundance $12 + \log(\text{O}/\text{H})$ vs. gas-phase metallicity (mass fraction of all metals) Z_{gas} . *Right:* Stellar iron abundance $[\text{Fe}/\text{H}]$ vs. stellar metallicity Z_* . Black dots collect all the data points presented in this work. The cyan lines represent the best fits of these relations with slope unity. These definitions give essentially identical results, and are equivalent, for all of our results in this paper.

epochs we present earlier in this paper. Both relations are extremely tight and have slope unity, which ensures the validity, at least to the first order, to use either quantity to represent metallicities interchangeably. The best fits for our simulations are $12 + \log(\text{O}/\text{H}) = \log(Z_{\text{gas}}/Z_{\odot}) + 9.0$ and $[\text{Fe}/\text{H}] = \log(Z_*/Z_{\odot}) - 0.20$. We emphasize that these relations may suffer from systematic uncertainties that originate from: (1) Type-II and Type-I SNe rates, (2) metal yields of tracked species from different channels, and (3) the solar abundances we adopt in our simulations.